

Unit 11 - Week 10: Data Compression-2 (Unit 10) and Data Compression-3 (Unit 11)

Course outline
How does an NPTEL online course work?
Week 1: Information and probabilistic modeling
Week 2: Uncertainty, compression, and entropy
Week 3: Randomness and entropy
Week 4: Information and statistical inference-1
Week 5: Information and statistical inference-2
Week 6: Properties of measures of Information-1
Week 7: Properties of measures of Information-2 and Information Theoretic lower bounds
Week 8: Information theoretic lower bounds and Data Compression-1
Week 9: Data Compression-1 (Unit 9) and Data Compression-2 (Unit 10)
Week 10: Data Compression-2 (Unit 10) and Data Compression-3 (Unit 11)
<input type="radio"/> Information Theory Review 9 <input type="radio"/> Arithmetic code <input checked="" type="radio"/> Online probability assignment <input checked="" type="radio"/> Compression of databases: A scheme <input type="radio"/> Compression of databases: A lower bound
<input type="radio"/> Quiz : Weekly Assignment 10 <input checked="" type="radio"/> Unit 10 Notes (contd.) <input type="radio"/> Solution 10
Week 11: Channel coding and capacity (Unit 12) and Shannon's channel coding theorem proof (Unit 13)
Week 12: Shannon's channel coding theorem proof (Unit 13) and Gaussian channels (Unit 14)
VIDEO DOWNLOAD
Live Session

Weekly Assignment 10

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2020-11-25, 23:59 IST.

1) Suppose $X^n = (X_1, \dots, X_n)$ is distributed as P^n , where $P = \text{Ber}(p)$. Let $K(X^n)$ denote the number of times '1' appears in X^n . Which of the following are true? 1 point

(Recall: P^n denotes the n -fold product distribution; $h(p) \triangleq -p \log p - (1-p) \log(1-p)$)

- $\mathbb{E}[K(X^n)] = p$
- $\mathbb{E}[K(X^n)] = np$
- $\mathbb{E}\left[h\left(\frac{K(X^n)}{n}\right)\right] \geq h\left(\frac{1}{n}\mathbb{E}[K(X^n)]\right)$
- $\mathbb{E}\left[h\left(\frac{K(X^n)}{n}\right)\right] \leq h(p)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\mathbb{E}[K(X^n)] = np$
 $\mathbb{E}\left[h\left(\frac{K(X^n)}{n}\right)\right] \leq h(p)$

2) Suppose $X^n = (X_1, \dots, X_n)$ is distributed as P^n , where $X_i \in \{1, \dots, k\}$ for every $i \in \{1, \dots, n\}$. Let $Q(X^n)$ denote the type of the sequence X^n . Which of the following are true? 1 point

($\mathbf{1}_{\{1\}}$ denotes an indicator random variable.)

- $Q(X^n) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=1\}}, \dots, \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=k\}}\right)$
- $Q(X^n) = \left(\frac{1}{n} \sum_{i=1}^k \mathbf{1}_{\{X_i=1\}}, \dots, \frac{1}{n} \sum_{i=1}^k \mathbf{1}_{\{X_i=n\}}\right)$
- $\mathbb{E}[Q(X^n)] = P$
- $\mathbb{E}[Q(X^n)] = (1/k, \dots, 1/k)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$Q(X^n) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=1\}}, \dots, \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=k\}}\right)$
 $\mathbb{E}[Q(X^n)] = P$

3) Suppose $X^n = (X_1, \dots, X_n)$ is distributed as P^n , where $X_i \in \mathcal{X}$ for every $i \in \{1, \dots, n\}$. Let $\bar{P}^{(n)}$ be the distribution of X^n conditioned on the event $\{X^n \in \mathcal{T}_R^{(n)}\}$, where $\mathcal{T}_R^{(n)} \subset \mathcal{X}^n$ is the set of sequences of type R . Which of the following are true? 1 point

- $\bar{P}^{(n)}(x^n) = 1/|\mathcal{X}|^n$ for every $x^n \in \mathcal{X}^n$
- $\bar{P}^{(n)}(x^n) = 1/|\mathcal{T}_R^{(n)}|$ for every $x^n \in \mathcal{X}^n$
- $\bar{P}^{(n)}(x^n) = 1/|\mathcal{T}_R^{(n)}|$ for every $x^n \in \mathcal{T}_R^{(n)}$
- $\bar{P}^{(n)}(x^n) = 0$ for every $x^n \notin \mathcal{T}_R^{(n)}$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\bar{P}^{(n)}(x^n) = 1/|\mathcal{T}_R^{(n)}|$ for every $x^n \in \mathcal{T}_R^{(n)}$
 $\bar{P}^{(n)}(x^n) = 0$ for every $x^n \notin \mathcal{T}_R^{(n)}$

4) Let $x \in [0, 1]$, and let $x = 0.b_1b_2\dots$ be a binary expansion of x (e.g. $1/4 = 0.01$). Let $[x]_\ell = 0.b_1b_2\dots b_\ell$. Which of the following are true? 1 point

- $x < [x]_\ell + 2^{-\ell-1}$
- $x < [x]_\ell + 2^{-\ell}$
- $x \geq [x]_\ell$
- $x \geq [x]_\ell + 2^{-\ell}$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$x < [x]_\ell + 2^{-\ell}$
 $x \geq [x]_\ell$

5) Let $[a, b) \subset [0, 1)$. Clearly, $c = \frac{a+3b}{4} \in [a, b)$; let $c = 0.b_1b_2\dots$ be its binary expansion. Let $[c]_\ell = 0.b_1b_2\dots b_\ell$. Which of the following are true? 1 point

- $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) \rceil$
- $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 1 \rceil$
- $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 2 \rceil$
- $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 3 \rceil$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 1 \rceil$
 $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 2 \rceil$
 $[c]_\ell \in [a, b)$ for $\ell = \lceil -\log(b-a) + 3 \rceil$

6) Suppose an i.i.d. source outputs symbols from alphabet $\{1, 2, 3, 4\}$ with probabilities $P = (1/3, 1/6, 1/6, 1/3)$. Recall that to encode n -length sequences using arithmetic coding, we assign an interval to each n -length sequence. Suppose $n = 4$. Which of the following are true? 1 point

- Sequence '3142' is assigned an interval of length $1/324$.
- Sequence '2314' is assigned an interval of length $1/324$.
- Sequence '1234' is assigned an interval of length $1/324$.
- Sequence '1111' is assigned an interval of length $1/81$.

No, the answer is incorrect.

Score: 0

Accepted Answers:

Sequence '3142' is assigned an interval of length $1/324$.
 Sequence '2314' is assigned an interval of length $1/324$.
 Sequence '1234' is assigned an interval of length $1/324$.
 Sequence '1111' is assigned an interval of length $1/81$.

7) Suppose \mathcal{F} is a family of functions from $\mathcal{X} \rightarrow \{1, \dots, m\}$. Let $x_1, \dots, x_n \in \mathcal{X}$. Let F be a function sampled uniformly at random from \mathcal{F} . For $j \in \{1, \dots, m\}$, let N_j be the number of symbols out of x_1, \dots, x_n that are mapped to j by F . That is, $N_j = \sum_{i=1}^n \mathbf{1}_{\{F(x_i)=j\}}$. Which of the following are true? 1 point

- $N_j^2 = \sum_{i=1}^n \mathbf{1}_{\{F(x_i)=j\}}^2$
- $N_j^2 = \sum_{i=1}^n \sum_{k=1}^n \mathbf{1}_{\{F(x_i)=j\}} \mathbf{1}_{\{F(x_k)=j\}}$
- $\sum_{j=1}^m N_j^2 = \sum_{i=1}^n \sum_{k=1}^n \mathbf{1}_{\{F(x_i)=F(x_k)\}}$
- $\sum_{j=1}^m N_j^2 = n + \sum_{i=1}^n \sum_{k=1}^n \mathbf{1}_{\{F(x_i) \neq F(x_k)\}}$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$N_j^2 = \sum_{i=1}^n \sum_{k=1}^n \mathbf{1}_{\{F(x_i)=j\}} \mathbf{1}_{\{F(x_k)=j\}}$
 $\sum_{j=1}^m N_j^2 = \sum_{i=1}^n \sum_{k=1}^n \mathbf{1}_{\{F(x_i)=F(x_k)\}}$

8) Consider the same set-up as in Q.6. Which of the following are true? 1 point

- $\mathbb{E}[N_j^2] = \sum_{i=1}^n \mathbb{P}(F(x_i) = j)^2$
- $\mathbb{E}[\sum_{j=1}^m N_j^2] = \sum_{i=1}^n \sum_{k=1}^n \mathbb{P}(F(x_i) = F(x_k))$
- $\mathbb{E}[\sum_{j=1}^m N_j^2] = n + \sum_{i=1}^n \sum_{k=1}^n \mathbb{P}(F(x_i) = F(x_k))$
- $\mathbb{E}[\sum_{j=1}^m N_j^2] = n + 2 \sum_{i=1}^n \sum_{k=i+1}^n \mathbb{P}(F(x_i) = F(x_k))$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\mathbb{E}[\sum_{j=1}^m N_j^2] = \sum_{i=1}^n \sum_{k=1}^n \mathbb{P}(F(x_i) = F(x_k))$
 $\mathbb{E}[\sum_{j=1}^m N_j^2] = n + 2 \sum_{i=1}^n \sum_{k=i+1}^n \mathbb{P}(F(x_i) = F(x_k))$

9) Let X_1, \dots, X_n be random variables with joint distribution P_{X^n} . Let $Y = g(X_1, \dots, X_n)$ where $g: \mathcal{X}^n \rightarrow \mathcal{Y}$ is a deterministic function. Further, let $Z = h(Y, U)$ where $h: \mathcal{Y} \times \{1, \dots, n\}$ is a deterministic function, and U is a $\text{Ber}(1/2)$ random variable independent of X^n . Which of the following are true? 1 point

($H(X|Y) = y$) denotes the entropy of the conditional distribution $P_{X|Y}(\cdot|y)$.)

- $H(X_Z|Y, Z) = \sum_{i=1}^n \mathbb{P}(Z = i)H(X_i|Y)$
- $H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Z = i) \mathbb{P}(Y = y|Z = i)H(X_i|Y = y)$
- $H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Y = y, Z = i)H(X_i|Y = y)$
- $H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Y = y, Z = i)H(X_i|Y = y, Z = i)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Z = i) \mathbb{P}(Y = y|Z = i)H(X_i|Y = y)$
 $H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Y = y, Z = i)H(X_i|Y = y)$
 $H(X_Z|Y, Z) = \sum_{y \in \mathcal{Y}} \sum_{i=1}^n \mathbb{P}(Y = y, Z = i)H(X_i|Y = y, Z = i)$

10) Let B_1, \dots, B_n be i.i.d. random variables with $B_i \in \{1, 2, 3\}$ for each $i \in \{1, \dots, n\}$. Let $M = g(B_1, \dots, B_n)$, where g is a deterministic function. Let $J \in \{1, \dots, n\}$ be sampled uniformly at random, independently of M . Let $P_e = \mathbb{P}(h(M, J) \neq B_J)$, where h is a deterministic function. Which of the following are true? 1 point

- $H(B_J|M, J) = \frac{1}{n} \sum_{i=1}^n H(B_i|M)$
- $H(B_J|M, J) = \log 3$
- $H(B_J|M, J) \leq P_e + h(P_e)$
- $H(B_J|M, J) \leq h(P_e)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$H(B_J|M, J) = \frac{1}{n} \sum_{i=1}^n H(B_i|M)$
 $H(B_J|M, J) \leq P_e + h(P_e)$